DOCUMENT RESUME

ED 204 353 TM 810 293

AUTHOR Borgen, Fred H.; Muchinsky, Paul M.
TITLE Convergence and Robustness of Analytic Techniques for Qualitative Data. Final Report, July 1, 1978 to September 30, 1979.
INSTITUTION Iowa State Univ. of Science and Technology, Ames.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.
PUB DATE Sep 79
GRANT NIE-G-78-0066
NOTE 23p.

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Statistical Analysis
IDENTIFIERS *Chi Square; Coefficient of Association; Contingency Tables; Monte Carlo Methods; Power (Statistics); *Qualitative Data; Robustness

ABSTRACT

The purposes of this study are to compare alternative statistical techniques for analysis of the kinds of qualitative data commonly obtained in educational settings. Distinctions between chi-square based measures for contingency tables and proportional reduction of error measures of the strength of association are reviewed, and it is argued that there are direct equivalencies between the chi-square based measures and at least one of the proportional reduction measures, Goodman and Kruskal's tau. A direct arithmetic linkage between tau, chi-square, and Cramer's V is demonstrated, and the implications of this linkage are discussed. In order to study the robustness of these various analytic techniques, the Monte Carlo simulation technique was used to construct 45 different contingency tables, varying in sample size and levels of Cramer's V. Power differences for chi-square and tau indicate that underlying relationships are many times more likely to be detected by chi-square than by tau. The final question of this study was the robustness of chi-square tests when the contingency table marginal totals were significantly skewed. The power decrement was about 10-15% when the skew was extreme; the adverse effect of moderate skew was quite mild. (BW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

ED204353

FINAL REPORT TO

NATIONAL INSTITUTE OF EDUCATION

CONVERGENCE AND ROBUSTNESS OF ANALYTIC TECHNIQUES FOR QUALITATIVE DATA

Fred H. Borgen and Paul M. Muchinsky

Iowa State University

Ames, Iowa

TM 810 293

NIE Small Grant NIEG780066
July 1, 1978 to September 30, 1979

INTRODUCTION AND BACKGROUND

Overview

Educational data frequently have qualitative forms, such as the attainment of criterion level on a mastery test. In attempting to study such performance, the most useful techniques are needed for relating such qualitative data to other variables. This study provides some guidelines about the kinds of questions and conditions under which alternative analytic techniques can be most effectively used. The purposes of this study are to compare alternative statistical techniques for analysis of the kinds of qualitative data commonly obtained in educational settings. Methods to be compared are measures of association for contingency tables, linear logistic models, and classical linear models such as discriminant analysis. The first major purpose is to compare these diverse techniques with respect to their convergence to similar analytic conclusions. The second major purpose is to evaluate the techniques with respect to robustness, that is, their accuracy when the data deviate from ideal conditions of sample size and skewness. Monte Carlo computer-simulation techniques are used to generate large "populations" of contingency table data with specified structures. Two variables are experimentally manipulated: degree of skewness in the population and sample size. Samples are drawn from these populations and successively analyzed by the major types of analytic techniques. The results from the different techniques are evaluated with respect to their convergence on similar conclusions and their robustness.

Background

Often in educational practice and research, important data are qualitative in form. For example, a student's score on a mastery test might be

expressed as reaching criterion or not reaching criterion. We might wish to study mastery performance in terms of its relationship to other variables. Such categorical data may not be tractable with classical statistical techniques designed for continuous data. However, unlike statistical techniques for continuous, quantitative data, we do not have equally extensive information about the variety of conditions under which techniques for qualitative data can be used with confidence. The intent of this study is to provide some comparative guidelines for such techniques.

Qualitative data. Qualitative data have long had a place in the repertoire of statistical techniques, at least as far back as Pearson's original work with chi-square. However, recent work has vastly extended the possibilities for analysis, and many recent sources have presented methods for use with qualitative data. Major general sources are Fleiss (1973), Hays (1973), Cox (1970), Haberman (1974), and Bishop, Fienberg, and Holland (1975).

Educational applications. One can identify particular educational problems where innovative techniques are being developed for qualitative data. One area is assessing agreement among nominal categories, such as ratings (Tinsley & Weiss, 1975). Another area is criterion-referenced, or mastery, testing where the person receives a pass- not pass dichotomous scores. Here new methods are being developed which will permit study of the reliability and validity of such data (Huynh, 1976; Subkoviak, 1976). Finally, the latent trait models of testing have an underlying qualitative assumption and their work has promise for further understanding of qualitative analysis (Rasch, 1960; Hambleton & Cook, 1977).

Monte Carlo Methods

Monte Carlo methods have played a major role in establishing practical guidelines for the applied user of statistical methods. Elegant theory specifies the properties and uses of specific statistical distributions when the theoretical assumptions are met. But in the real world of data analysis it is unlikely that the assumptions of statistical theory will be met perfectly. Thus, the real question for the applied use of statistics is not, "Have I met the assumptions of the statistical model," but rather, "How much may I deviate from the assumptions without seriously distorting the inferences I wish to make" (Glass et al., 1972). Thus, the practical issue for statistics becomes robustness, the quality of statistical inferences possible when assumptions are not met. Often it is not feasible to mathematically evaluate robustness because of the complexity of the issues. Consequently robustness has most frequently been evaluated by the Monte Carlo method, which entails computer-generated samples of data with specified parameters to simulate data differing from normal-theory assumptions in specified and varied ways. Such Monte Carlo data, with known population values, are then analyzed with a particular statistical test to determine how closely the test can detect the known relationships in the data.

There now is an established technology and literature for the application of Monte Carlo methods (Hammersley & Handscomb, 1965; Sobol, 1974; Wherry et al., 1965; Montanelli, 1975; Hildebrand et al., 1977). Researchers in education, psychology, and sociology have particularly been concerned about the quality of their data for meeting the stringent assumptions of most parametric statistical models. One result of this concern was a premature rush during the 1950s to "nonparametric," "distribution-free" methods. Another approach, which ultimately has been more productive, has been to

evaluate the practical effects of "violated" assumptions with parametric statistics using Monte Carlo methods (Collier & Larson, 1969; Glass et al., 1972; Urry, 1970; Schmitt et al., 1977). A series of studies evaluating the robustness of common statistics have been conducted, many of which are cited in the section which follows.

Robustness of Statistical Tests

As mentioned previously, the applied researcher is appropriately concerned about the robustness of a particular statistical test, i.e., the extent to which the test permits accurate inference when the formal assumptions of the test are violated in varying degrees. Applied researchers in the behavioral sciences, in particular, have been examining these issues at least since the early 1950s (Box, 1954; Lindquist, 1953), with much of the earliest work examining the robustness of the F-test. Monte Carlo studies of robustness are continuing today, recently stimulated by the availability of large-scale, rapid computing power for the generation of large and complexly defined samples. Hence, the more recent studies have been able to examine several parameters describing samples, and also to evaluate the interactive effects of several parameters on robustness.

The general results of Monte Carlo studies have been to show that the more commonly used "parametric" statistical tests are moderately to extremely robust over a large range of sampling deviation from the theoretical assumptions. Additionally, the larger studies have been able to pinpoint the configurations of parameters where sampling inference is likely to be in error. These studies have focused on statistical tests appropriate for continuous, quantitative variables; lesser work is visible in the literature on methods appropriate for qualitative or nominal data. Among the former kinds of tests,

robustness has been demonstrated for t-tests (Boneau, 1959; Havlicek & Peterson, (1974), for F-tests and analysis of variance (Boneau, 1959; Scheffe', 1959; Lunney, 1970; Mandeville, 1972; Glass et al., 1972), for correlations with Pearsonian r (Zeller & Levine, 1974; Havlicek & Peterson, 1977), and for multiple regression (Bohrnstedt & Carter, 1971; Dawes & Corrigan, 1974). Thus while much is known about the robustness of statistical methods for continuous, quantitative data, there is apparently less known about the robustness of appropriate statistics for qualitative data which may be in such forms as frequencies, proportions, classes, etc.

CONVERGENCE RESULTS

Chi-Square and Measures of Association For Contingency Tables

Ever since the classic exposition of Goodman and Kruskal (1954), the standard statistical texts (e.g., Hays, 1973; Bishop, Fienberg, and Holland, 1975) have outlined the distinction between chi-square based measures for contingency tables and proportional reduction of error measures of the strength of association. Specifically, Goodman and Kruskal (1954) introduced proportional reduction in error statistics such as lambda, gamma, and tau. These statistics were seen as conceptually superior to the chi-square based methods because they have greater interpretability. In this section, however, we attempt to show that the distinction made by Goodman and Kruskal is overdrawn, and in fact there are direct equivalancies between the chi-square based measures and at least one of the proportional reduction of error measures, namely tau. These equivalences hold under certain specified conditions which have considerable generality. The equivalences lead to exploiting the conceptual advantages of both chi-square and tau in the same situations. Specifically, it is possible with tau to make nonzero point estimates of

the strength of association; and similarly with chi-square to test against the hypothesis of no association. Moreover, it is possible to have the interpretive advantages concurrently or alternatively of the two approaches. In other words, the approaches have different conceptual bases, and yet are mathematically equivalent, therefore, the conceptual interpretations can in many cases be treated as equivalent. These conceptual and arithmetic integrations lead to the possibilities for interpreting chi-square in a contingency table as a measure of proportional reduction of error in prediction. To the best of our knowledge, this result is first obtained here.

Following are some of the linkages between chi-square, tau, Cramer's V, and Margolin and Light's (1974) or Light and Margolin's C^2 statistic (referred to as U^2 in Bishop et al, 1975). Our attention was initially directed to these equivalencies by our Monte Carlo studies with a specified sample size and table size, wherein there was a perfect (1.0) correlation between chi-square and Goodman and Kruskal's tau. This led to an examination of algebraic equivalences. After some trial and error, it became apparent that tau could be specified as a ratio of obtained chi-square to maximum possible chi-square for a given table. Further, it became evident that this formula is very much like Cramer's V. In fact, this formula for the largest tau in a table (from the two asymmetric measures) is equivalent to Cramer's V squared. Examination of the detailed algebraic computational formulas for the two statistics also show this equivalence

$$\text{Tau (in 2 X J Case)} = \frac{\text{Obt } \chi^2}{\text{Max } \chi^2} = \frac{\chi^2}{N}$$

$$\text{Cramer's } v = \sqrt{\frac{\chi^2}{N}} \qquad \text{tau} = v^2$$

$$\chi^2 = 16 \quad \text{for} \quad v = .4 \qquad N = 100$$

$$v = \sqrt{\frac{16}{100}} = \sqrt{.16} = .4$$

$$\text{tau} = .16$$

The implications of this direct arithmetic linkage between tau, chi-square, and Cramer's V are several-fold. Most far reaching perhaps, is the implication that chi-square can be interpreted in a strength of association without looking only at zero relationships. Specifically, Goodman and Kruskal (1972) present the asymptotic variance for tau. This approach then, because of the above linkages, should be directly applicable to testing nonzero values of chi-square. Also, it makes sense to directly test Tau against a null relationship by resorting to chi-square. (Later it is shown that Margolin and Light have attempted to do this with C^2; we believe however, that we can show, and do show a more parsimonious interpretation directly with chi-square.) In fact, we argue that the C^2 statistic is in practical terms superfluous and that chi-square alone is sufficient for that purpose. Nonetheless, the Margolin and Light (1974, 1971) contributions for phrasing C^2 in analysis of variance terms are conceptually potent.

We have independently discovered that tau and chi-square are closely linked. After this naive discovery, forays to the literature have revealed to us some published confirmations of this result for certain specified cases. For example, Goodman and Kruskal (1954) in introducing tau indicated that it takes a "chi-square like" form, however, they do not state that it <u>is</u> equivalent to chi-square, and apparently (personal communication with W. Kruskal, in May 1979) have never considered it equivalent to chi-square. Additionally, Margolin and Light (1974) present results arguing that tau is equivalent to a function of phi-squared, but only under limited conditions of I=2, i.e., dichotomous response variables. This result is repeated in the comprehensive text of Bishop et al. (1975). However, we argue that this equivalence between tau and the chi-square statistics is much more general than for the dichotomous response case, and can be generalized quite broadly.

The Margolin and Light (1974) article seems to us quite profound in its implications. They present an analysis of variance model for chi-square, drawing from the early work of Gini (1912). Their derivations also lead them to demonstrate that tau is equivalent to the ratio of BSS/TSS. This apparent capacity to phrase categorical data in analysis of variance terms, and hence within the general linear model, seems to have fairly wide possibilities for both theory and application. We do not attempt here to exploit all of these possibilities.

Margolin and Light (1971; 1974) introduce a statistic called C^2 which permits a testing of tau-b against a null population relationship. This statistic seems totally equivalent to our derivation of the relationship of tau to chi-square with the exception that we use the multiplier n, whereas Margolin and Light use the multiplier n-1. They argue that C^2 is approximated by a chi-square distribution. Superficial thinking suggests that C^2 should be exactly chi-square when the N multiplier is used.

An intriguing problem is the character of larger tau-b. Tau is an asymmetric statistic, as is well known and introduced at the outset in Goodman and Kruskal's 1954 article. Our investigations demonstrate the strict arithmetical relationships between larger tau and chi-square based statistics. We have not at this time been able to show direct functional relationships for the smaller tau value occurring in most contingency tables. Larger tau-b always occurs when one is predicting from the larger number of categories to the smaller number.

One intriguing side issue is the report of correlations among measures given by Light and Margolin (1971) and Margolin and Light (1974). In no cases do they report 1.0 correlations between tau and the chi-square based measures. Our investigations lead us to expect that under some conditions

perfect correlations would be obtained here. It is not immediately apparent why they did not obtain perfect correlations.

MONTE CARLO DATA GENERATION

The simulation task is to construct contingency tables having the desired characteristics in the population. A first conceptual approach was to think of a table with a very large sample size, say 1,000,000, and then to randomly draw from this large table to form a specific table with small n. Some original experimentation was done with this approach before we discovered a much more efficient procedure. This conceptual approach recognizes that for a given table with specified skew and strength of association, the proportion of frequencies with each cell is fixed. Thus, for example, for a 2 x 3 table with .5 skew (equal marginal totals on the dichotomous variable) and a strength of association shown by a Cramer's v of .10, the proportion of frequencies in each cell would be:

.146254	.187079	.3333
.166667	.166667	.3333
.187079	.146254	.3333
.500	.500	1.000

We derived these values with the following rationale:

$$\text{Cramer's } v = \sqrt{\frac{\chi^2}{N}}, \text{ where N = sample size}$$

For a desired Cramer's of .10 and an N of 100, the corresponding χ^2 can thus be solved from the above as 1.0.

The 4 corners of the 2 x 3 table contribute equally to the χ^2 of 1.0. Thus, each corner of the table must have a χ^2 of .25.

The χ^2 of .25 is solved thus:

$$\chi^2 = .25 = \frac{(\text{Observed} - \text{Expected})^2}{\text{Expected}}$$

For this case the expected values are 16.67, so we can solve for the observed value for one corner cell. Filling in the remainder of the table is then simple arithmetic.

The upper borders of the cells of the table in our example then can be demarcated with cumulative proportions:

.146254	.333333
.500000	.666667
.853745	1.00000

An IBM system random-number generator producing random numbers evenly distributed between 0.0 and 1.0 is activated for each frequency in a table. Each number is tested against the above cumulative proportions and where a match occurs the count for the appropriate cell is incremented by 1. The appendix gives more detail about the FORTRAN programming for this routine.

ROBUSTNESS RESULTS

Power of Chi-square for 2 x 2 Contingency Tables

Using the Monte Carlo simulation technique outlined above, 2 x 2 contingency tables were generated with samples sizes of 20 to 100, in increments of 10. The simulation specified five non-zero levels of association in the underlying theoretical population. Measuring strength of association with Cramer's v, the five levels were .1, .2, .3, .4 and .5. Thus, with 9 sample sizes and 5 levels of Cramer's v, a total of 45 different kinds of 2 x 2 tables were simulated. For each of these conditions, 1000 separate contingency tables (1000 replications) were randomly generated.

Figure 1 shows results for the 2 x 2 tables in terms of the power of the chi-square test, i.e., the probability of detecting with a significant statistical result a true population relationship for given conditions of sample size and underlying relationship. The table shows quite orderly curves, since each point represents the proportion of times chi-square was significant over 1000 replications. The results here are congruent with those provided by Cohen (1977), although these results are presented for a greater number of conditions. The figure shows that power (using the .01 significance level) for the stronger relationships (v = .4 or .5) goes over 50% when the sample size reaches about 50. Conversely, when the underlying true relationship is modest (v = .1, .2 or .3), the chi-square test does not have a high likelihood of detecting it at .01 significance level when samples are 50 or less in size. Moreover, when the true relationship is mild (v = .1), the chance of detecting it with a significant chi-square (.01 level) is no better than 5%, even for samples as large as 100. The curve for v = .1 is nearly flat, yielding the practical result that increases in sample size produce little gain in statistical power. As the strength of association increases the power curves become progressively steeper, meaning that with greater underlying association, power is more affected by increasing sample size.

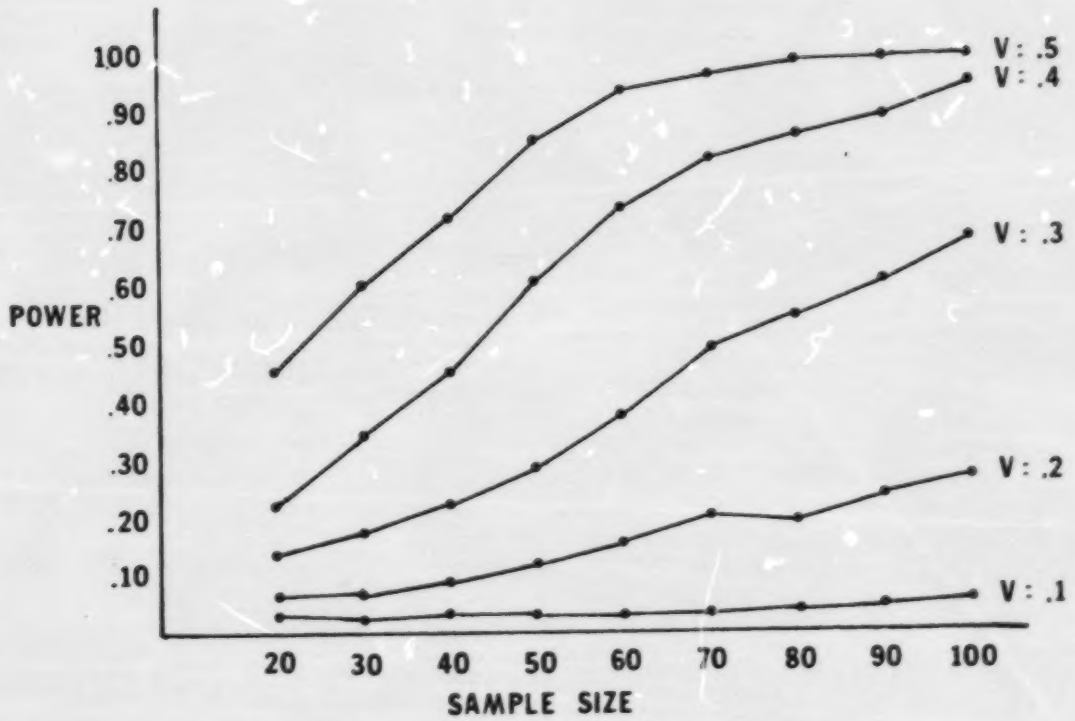

Fig. 1. Power of chi-square for 2 x 2 contingency tables. (Proportion of chi-square tests significant at .01 level for specified Cramer's V, the strength of association; each point based on 1000 replications.)

Comparative Power of Chi-square and Tau for 2 x 2 Contingency Table

The simulation described immediately above was also used to evaluate the power of Goodman and Kruskal's (1963) asymmetric tau statistic. Standard error for tau was initially calculated using the formula provided by Bishop et al. (1975). Our work with this formula suggested it was in error; following phone conversation with Fienberg, he concluded there was indeed an error in the formula. Subsequently, the tau standard error was calculated with the corrections provided by Fienberg.

Figure 2 shows comparative power results for chi-square and tau when the underlying relationships are at the levels of .4 and .5 v. Significance level chosen for this power chart is .05 because of the very minimum power exhibited by tau at the .01 level. The power differences for chi-square and tau are striking. For nearly all conditions, underlying relationships are many times more likely to be detected by chi-square than by tau. Some serious limitation remains in the derivation and computation of the standard error for tau.

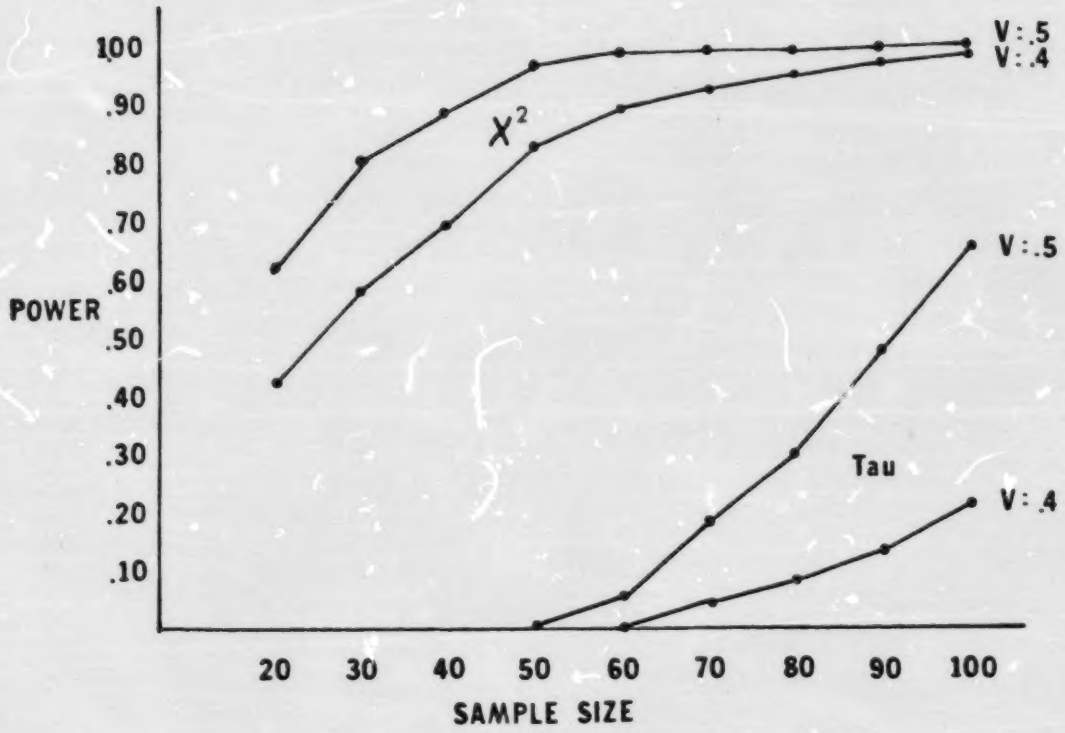

Fig. 2. Comparison of power for chi-square and tau for 2 x 2 contingency tables. (Proportion of tests significant at .05 level for two levels of Cramer's V, the strength of association; each point based on 1000 replications.)

Effects of Skew and Sample Size on Chi-Square Power: 2 x 3 Tables

A prime question of this study was the robustness of chi-square tests when the contingency table marginal totals were significantly skewed. We simulated tables with marginals for the dichotomous variable being proportionally .1 and .9 (.1 skew); .3 and .7 (.3 skew); and .5 and .5 (.5 skew). Tables were constructed with sample size varying from 10 to 100, in increments of 10. With three levels of skew and 10 sample sizes there were a total of 30 table configurations. Cramer's v for each case was set at .4. For each of the 30 tables, 250 replications were run.

Figure 3 shows the power of chi-square for the combinations of skew and sample size. The significance level was set at .01. Due to the lower number of replications here (250 versus 1000), these power curves are not as smooth as those shown in Figures 1 and 2. (The larger number of replications is, of course, preferable when computer costs permit.) The point on the curve for skew .1, sample size 20, appears to be somewhat high, as an artifact of random sampling error. The curves show some adverse effect of skew on power when the sample sizes are thirty or below. The power decrement, for these small samples, appears to be about 10 to 15% when the skew is extreme. The adverse effect of moderate skew is quite mild. To our knowledge, this is the first presentation of the adverse consequences of skew on chi-square power. It is a result worthy of additional exploration, including somewhat more precise specification using a larger number of replications. For larger samples, say 40 and above, there is very little adverse effect of skew, and the user can be as confident as usual using chi-square when skew is present.

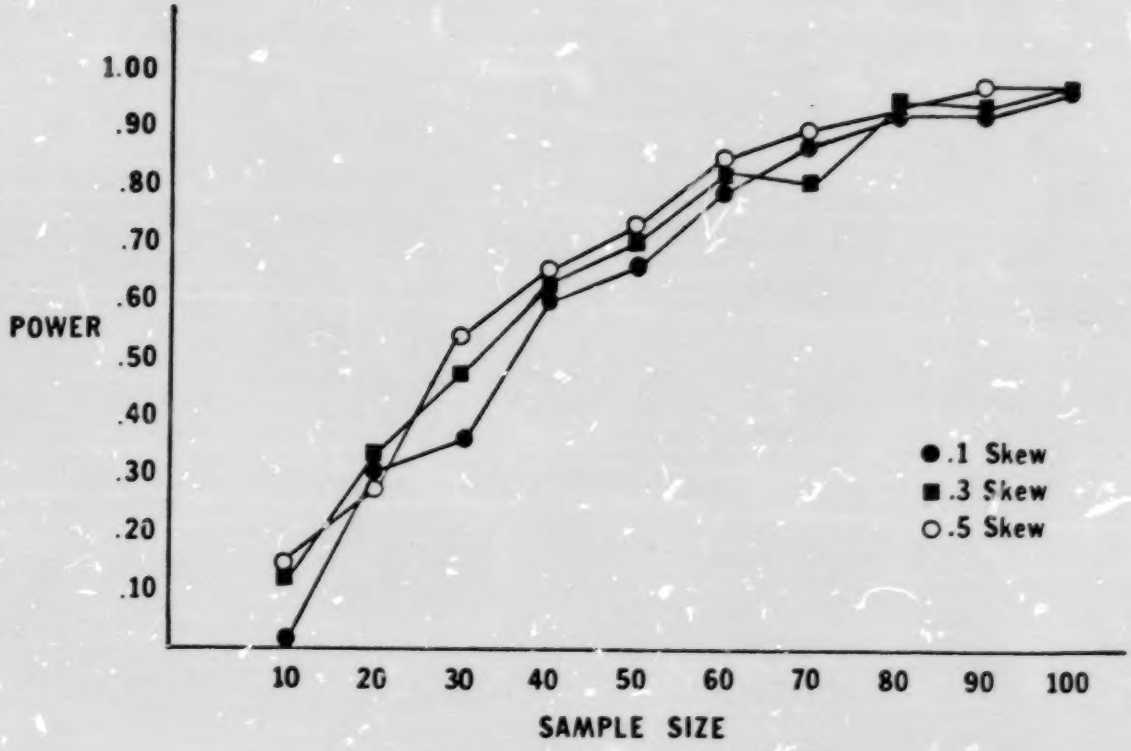

Fig. 3. Power of chi-square for specified levels of skew in 2 x 3 contingency tables. (Proportion of chi-square tests significant at .01 level; Cramer's V, the strength of association is .4; each point based on 250 replications.)

REFERENCES

Bishop, Y. M. M., Fienberg, S. E. & Holland, P. W. Discrete multivariate analysis: Theory and practice. Cambridge, Mass: MIT Press, 1975.

Bohrnstedt, G. W. & Carter, T. M. Robustness in regression analysis. In: Costner, H. L. (Ed.), Sociological Methodology 1971. San Francisco: Jossey-Bass, 1971.

Boneau, C. A. The effects of violations of assumptions underlying the t test. Psychological Bulletin, 1959, 57, 49-64.

Box, G E. P. Some theories on quadratic forms applied in the study of analysis of variance problems, I: Effects of inequality of variance in the one-way classification. Annals of Mathematical Statistics, 1954, 25, 290-302.

Cohen, J. Statistical power analyses for the behavioral sciences. New York: Academic Press, 1977.

Collier, R. O. Jr., & Larson, R. C. Monte Carlo methods. Review of Educational Research, 1969, 39, 734-747.

Cox, D. R. The analysis of binary data. London: Methuen, 1970.

Dawes, R. B. & Corrigan, B. Linear models in decision making. Psychological Bulletin, 1974, 81, 95-106.

Fleiss, J. L. Statistical methods for rates and proportions. New York: Wiley, 1973.

Gini, C. E. "Variability and Mutability", contribution to the study of statistical distributions and relations. Studi Economico-Giuridici della R. Universita de Cagliari, 1912.

Glass, G. V., Peckham, P. D. & Sanders, J. R. Consequences of failure to meet assumptions underlying the analysis of variance and covariance. Review of Educational Research, 1972, 42, 237-288.

Goodman, L. A. & Kruskal, W. H. Measures of association for cross-classifications. Journal of the American Statistical Association, 1954, 49, 732-764.

Goodman, L. A. & Kruskal, W. H. Measures of association for cross-classifications, IV. Simplification of asymptotic variances. Journal of the American Statistical Association, 1972, 67, 415-421.

Haberman, S. J. The analysis of frequency data. Chicago: University of Chicago Press, 1974.

Hambleton, R. K. & Cook, L. L. Latent trait models and their use in the analysis of educational test data. Journal of Educational Measurement, 1977, 14, 75-96.

Hammersley, J. M. & Handscomb, D. C. Monte Carlo Method. New York: Wiley, 1965.

Havlicek, L. L. & Peterson, N. P. Robustness of the t test: A guide for researchers on effect of violation of assumptions. Psychological Reports, 1974, 34, 1095-1114.

Havlicek, L. L. & Peterson, N. L. Effect of the violation of assumptions upon significance levels of the Pearson r. Psychological Bulletin, 1977, 84, 373-377.

Hays, W. L. Statistics for social scientists. New York: Holt, Rinehart & Winston, 1973.

Hildebrand, D. K., Laing, J. D. & Rosenthal, H. Prediction analysis of cross classifications. New York: John Wiley & Sons, 1977.

Huynh, H. On the reliability of decisions in domain-referenced testing. Journal of Educational Measurement, 1976, 13, 253-264.

Light, R. J. & Margolin, B. H. An analysis of variance for categorical data. Journal of the American Statistical Association, 1971, 66, 534-544.

Lindquist, E. F. The design and analysis of experiments in education and psychology. Boston: Houghton, Mifflin, 1953.

Mandeville, G. K. A new look at treatment differences. American Educational Research Journal, 1972, 9, 311-321.

Margolin, B. H. & Light, R. J. An analysis of variance for categorical data, II: Small sample comparisons with chi square and other competitors. Journal of the American Statistical Association, 1974, 69, 755-764.

Montanelli, R. G. Jr. A computer program to generate sample correlation and covariance matrices. Educational and Psychological Measurement, 1975, 35, 195-197.

Rasch, G. Probabilistic models for some intelligence and attainment tests. Copenhagen: Nielson and Lydiche, 1970.

Scheffe, H. The analysis of variance. New York: Wiley, 1959.

Schmitt, N., Coyle, B. W. & Rauschenberger, D. A Monte Carlo evaluation of three formula estimates of cross-validated multiple correlation. Psychological Bulletin, 1977, 84, 751-758.

Subkoviak, M. J. Estimating reliability from a single administration of a criterion-referenced test. Journal of Educational Measurement, 1976, 13, 265-276.

Tinsley, H. E. & Weiss, D. Interrater reliability and agreement of subjective judgments. Journal of Consulting Psychology, 1975, 22, 248-276.

Urry, V. W. A Monte Carlo investigation of logistic test models. Unpublished doctoral dissertation, Purdue University, 1970.

Wherry, R. J. Sr., Naylor, J. C., Wherry, R. J. Jr. & Fallis, R. F. Generating multiple samples of multivariate data with arbitrary population parameters. Psychometrika, 1965, 30, 303-313.

Zeller, R. A. & Levine, Z. H. The effects of violating the normality assumptions underlying r. Sociological Methods and Research, 1974, 2, 511-519.

APPENDIX

CHII: A Fortran Computer Program Implementing an Arithmetic Strategy for Monte Carlo Generation of Contingency Table Data

Description:

CHII is a routine for deriving contingency table samples from exactly specified populations via arithmetic procedures. Beginning with a fixed-point seed value, a floating-point, uniformly distributed variate with range 0-1.0 is first generated with a CALL to a library function. Then, depending on user specified parameters of table size (e.g., 2 X 3), sample size, and theoretical population distribution, the random variate is used in a series of test statements to make a determination of where in the table a single "one" should be entered. This process repeats itself a number of times equal to the sample size.

This procedure differs from typical Monte Carlo sampling method of first randomly creating a population, and the secondly sampling from it. Instead, CHII employs a strategy of checking each random variate's value against proportion values (user supplied, based on desired population characteristics), and then indexes arrays to keep track of the results. Owing to the recursiveness of all computerized pseudorandom variate generators, it is not necessary to create a population and subsequently sample from it: Samples may be created as they are needed for use. Testing of this routine, with sample sizes of 30, 60 and 90 in 2 X 3 tables, with population skews of .1, .3 and .5 reveals that the routine performs precisely as required. The prime advantage of the CHII strategy is tremendously reduced cost. Four-hundred and fifty samples of various sizes and theoretical skews required less than 10 sec to generate and output on an IBM 365.

Availability:

The routine is coded in FORTRAN IV in form suitable for IBM machines. It is short and easily adapted to other brands of computers. In the interest of programming efficiency users will change FORTRAN statements for their specific problems. Thus, the available documentation consists of a listing of CHII, sample output from a problem, and a short manual describing the general capabilities of this approach.